

Analyzing the Communication Clusters in Datacenters

Klaus-Tycho Foerster, Thibault Marette, Stefan Neumann,
Claudia Plant, Ylli Sadikaj, Stefan Schmid, Yllka Velaj



Overview

Context and motivation

- A lot of attention on datacenter traffic characteristics
- Most papers today mainly focus on how sparse and skewed^a, and exhibit locality^b
- Our work focuses on *clusters* in datacenter traffic matrix.

^a*ProjectToR Dataset*.

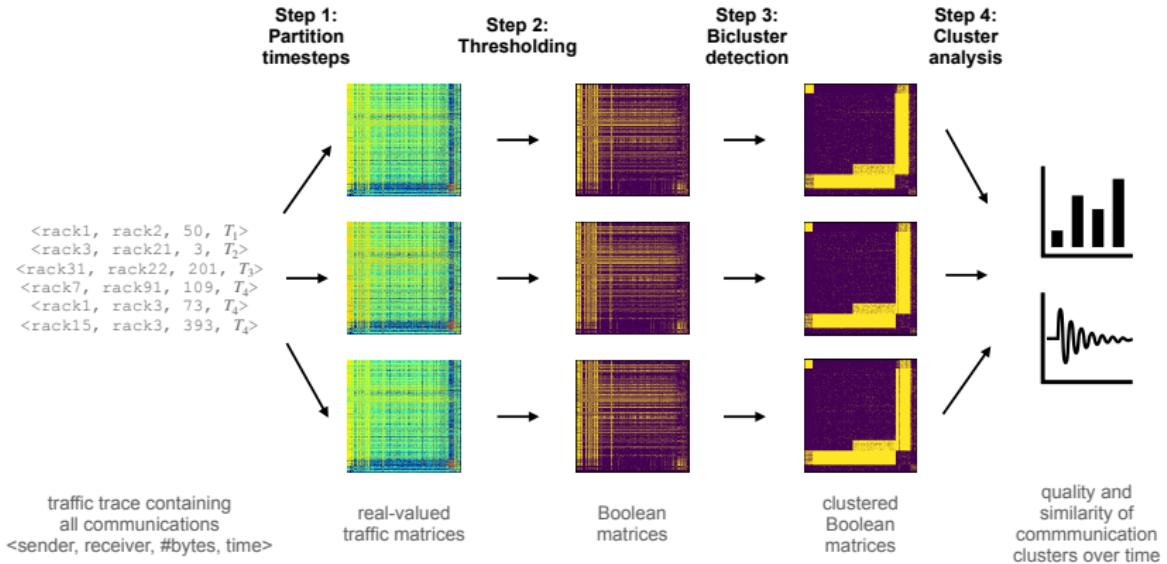
www.microsoft.com/en-us/research/project/projector-agile-reconfigurable-data-center-interconnect. 2016.

^bKai Chen et al. “OSA: An Optical Switching Architecture for Data Center Networks With Unprecedented Flexibility”. In: *IEEE/ACM Trans. Netw.* 22.2 (2014), pp. 498–511.

Focus of our paper

- Systematic and efficient approach to identify dense clusters
- Analyse quality and stability of found clusters

General context



An overview of our approach.

Experiments

Metrics

- Cluster quality using *precision* and *recall*
- Cluster relevance using *traffic inside biclusterings*
- Stability of clusters over time

Case study

Used Altoona datacenter, a real-world dataset containing actual datacenter packet traces released by facebook



Altoona datacenter

Results

Key takeaways

- Our approach is efficient and can find high-quality clusters
- We uncover that small clusters are responsible for a significant amount of network traffic
- These clusters are stable over time
- Our methodology only requires a list of endpoint pairs which cause a lot of traffic.
- Empirically that this gives good clusterings even when taking into account the absolute amounts of traffic

Make sure to checkout our article!